

What is claimed is:

1 1. A method of analyzing a target nucleic acid,
2 comprising:

3 (a) providing an array of probes comprising a probe
4 set comprising probes complementary to a reference sequence;

5 (b) hybridizing the target nucleic acid to the
6 array of probes;

7 (c) determining the relative hybridization of the
8 probes to the target nucleic acid,

9 (d) estimating the sequence of the target nucleic
10 acid from the relative hybridization of the probes;

11 (e) providing a further array of probes comprising
12 a probe set comprising probes complementary to the estimated
13 sequence of the target nucleic acid;

14 (f) hybridizing the target nucleic acid to the
15 further array of probes;

16 (g) determining the relative hybridization of the
17 probes to the target nucleic acid;

18 (h) reestimating the sequence of the target nucleic
19 acid from the relative hybridization of the probes.

1 2. The method of claim 1, further comprising
2 repeating steps (e)-(h) as necessary until the reestimated
3 sequence of the target nucleic acid is the true sequence of
4 the target nucleic acid.

1 3. The method of claim 1, wherein the target
2 nucleic acid is a species variant of the reference sequence.

1 4. The method of claim 1, wherein the reference
2 sequence is from a human and the target nucleic acid is from a
3 primate.

1 5. The method of claim 1, wherein the target
2 nucleic acid shows 50-99% sequence identity with the reference
3 sequence.

1 6. The method of claim 1, wherein the target
2 nucleic acid shows 80-95% sequence identity with the reference
3 sequence.

1 7. The method of claim 1, wherein the reference
2 sequence is at least 1000 nucleotides long, the array
3 comprises a probe set comprising overlapping probes that are
4 perfectly complementary to and span the reference sequence,
5 and the further array comprises probes that are perfectly
6 complementary to and span the estimated sequence.

1 8. The method of claim 1, wherein an estimated
2 sequence of the target nucleic acid includes a position of
3 ambiguity and the probe set showing perfect complementarity to
4 the estimated sequence includes a probe having including a
5 pooled nucleotide aligned with the position of ambiguity in
6 the target sequence.

1 9. The method of claim 1, wherein the reference
2 sequence is at least 10 kb.

1 10. The method of claim 1, wherein the reference
2 sequence is at least 1000 kb.

1 11. The method of claim 1, wherein the reference
2 sequence includes at least 90% of the human genome.

1 12. The method of claim 1, wherein the array of
2 probes comprises:

3 (1) a first probe set comprising a plurality of
4 probes, each probe comprising a segment of at least six
5 nucleotides exactly complementary to a subsequence of the
6 reference sequence, the segment including at least one
7 interrogation position complementary to a corresponding
8 nucleotide in the reference sequence,

9 (2) second, third and fourth probe sets, each
10 comprising a corresponding probe for each probe in the first
11 probe set, the probes in the second, third and fourth probe

12 sets being identical to a sequence comprising the
13 corresponding probe from the first probe set or a subsequence
14 of at least six nucleotides thereof that includes the at least
15 one interrogation position, except that the at least one
16 interrogation position is occupied by a different nucleotide
17 in each of the four corresponding probes from the four probe
18 sets.

1 13. The method of claim 12, wherein the sequence of
2 the target nucleic acid is estimated by:

3 (a) comparing the relative specific binding of four
4 corresponding probes from the first, second, third and fourth
5 probe sets;

6 (b) assigning a nucleotide in the sequence of the
7 target nucleic acid as the complement of the interrogation
8 position of the probe having the greatest specific binding;

9 (c) repeating (a) and (b) until each nucleotide of
10 interest in the sequence of the target nucleic acid has been
11 estimated.

1 14. The method of claim 1, wherein the sequence of
2 the target nucleic acid differs from the reference by at least
3 two positions within a probe length.

1 15. A method of analyzing a target nucleic acid,
2 comprising:

3 (a) designing an array of probes to be
4 complementary to an estimated sequence of the target nucleic
5 acid,

6 (b) hybridizing the array of probes to the target
7 nucleic acid;

8 (c) determining a reestimated sequence of the
9 target nucleic acid from the hybridization pattern of the
10 array to the target nucleic acid sequence to; and

11 (d) repeating (a)-(c) at least once.